2010. The Journal of Arachnology 38:377–382

SHORT COMMUNICATION

The utility of ITS2 in spider phylogenetics: notes on prior work and an example from *Anelosimus*

Ingi Agnarsson: Department of Biology, University of Puerto Rico - Rio Piedras (UPRRP), San Juan, PR, 00931-3360, Puerto Rico, USA.; Department of Entomology, National Museum of Natural History, Smithsonian Institution, NHB-105, PO Box 37012, Washington, D.C. 20013-7012, USA. E-mail: iagnarsson@gmail.com

Abstract. The ribosomal internal transcribed spacer ITS2 is probably the most popular nuclear DNA marker used to examine relationships among and within species in animals and plants. ITS2 sequences have also begun to be used as DNA barcodes. ITS2, however, has rarely been used in studies of spiders. Here, I examine the potential utility of this marker for spider phylogenetics based on preliminary data for *Anelosimus* spiders and a brief summary of prior work. The secondary structure of ITS2 facilitated alignment of highly divergent sequences and indicated that secondary structure morphology might be phylogenetically informative in itself. Phylogenetic analysis of *Anelosimus* species was congruent with a prior study based on a combination of six mitochondrial and nuclear loci plus morphology regarding the deeper clades within the genus. However, ITS2 had insufficient variation to resolve relationships within species and among closely related species. Previous studies have also discovered relatively little within-species variation in ITS2. In sum, ITS2 is an easily amplified and sequenced marker that is underutilized in spider phylogenetics; however, it has limited uses at the lowest taxonomic levels and is not likely to be a universally useful DNA barcode marker.

Keywords: Phylogeny, DNA barcode, ITS2 secondary structure, sociality, Theridiidae

ITS2, which is flanked by the 5.8S and large ribosomal subunit (28S) nuclear genes, is perhaps the most popular marker used to resolve relationships among and within species in animals and plants (Alverez & Wendel 2003; Bailey et al. 2003; Young & Coleman 2004; Schultz et al. 2005; Coleman 2009; Schultz & Wolf 2009). ITS2 sequences have also been proposed as effective DNA barcodes (e.g., Ben-David et al. 2007; Park et al. 2007). The popularity of this marker stems from a generally high level of variation, yet relatively conserved secondary structure, and ease of amplification and sequencing. However, comparatively few studies on spider phylogenetics have utilized this marker despite these benefits and a general paucity of good primers for nuclear markers. Among the few ITS2 studies in spiders, most focus on low taxonomic levels, reconstructing relationships among, and in some cases, within, species (Hedin 1997; Hormiga et al. 2003; Arnedo & Gillespie 2006; Chang et al. 2007; Bond & Stockman 2008). In spiders, ITS2 has generally been found to be a useful marker offering resolution at the species level, especially so in more genetically structured systems such as in trapdoor spiders (Bond & Stockman 2008), cave dwelling nesticids (Hedin 1997), and island radiations (Hormiga et al. 2003; Arnedo & Gillespie 2006). Other studies have used ITS2 as a tool to help separate closely related species. Variation allowing separation of closely related species/populations was found in *Poltys* (Smith 2006), *Pardosa* (Chang et al. 2007) and *Latrodectus* (Vink et al. 2008). However, variation was insufficient to separate closely related North American *Latrodectus* species (Zhang et al. 2004) or populations of *L. katipo* Powell 1870 (Vink et al. 2008).

This note reports on the utility of ITS2 data to resolve phylogenetic relationships among and within *Anelosimus* spider species, well known for their multiple origin of social behavior (Avilés 1997; Agnarsson 2006; Agnarsson et al. 2006). I use exemplar species from across the phylogeny of the genus and specimens representing most of the known 16S mitochondrial haplotype diversity within one species, *A. eximius* (Keyserling 1884). As the goal of this paper is practical application, I do not see a reason to prune the analyzed matrix to the exact ITS2 sequences, but I refer rather loosely to the entire region amplified by the FITS and RITS primers (see below) as ITS2.

I collected specimens in the field and placed them in 95% ethanol. Genitalia were abscised and stored as vouchers at the Zoological Museum of the University of Puerto Rico, while DNA was isolated from each individual using the prosoma, the abdomen, or both, with the QIAGEN DNAeasy Tissue Kit (Qiagen, Inc., Valencia, CA). I used the ITS-5.8S (FITS) and ITS-28S (RITS) primers (White et al. 1990) (FITS GGGACGATGAAGAACGGAGC, RITS TCCTCCGCTTATTGATATGC), using standard protocols with an annealing temperature of 47° C for 30 cycles. The PCR products were sequenced by the MACROGEN service, and sequences were submitted to GenBank (Accession numbers: HM584843–HM584883). Data for outgroups (*Latrodectus*, *Enoplognatha*), were obtained from GenBank. Preliminary alignments were done using ClustalW (Thompson et al. 1994) with gap opening and extension costs set at 24/6 and 8/2. Most of the sequences aligned readily; however, these preliminary alignments revealed an area of a particularly difficult alignment. Analyses of Clustal aligned matrices gave results largely incongruent with prior phylogenetic hypotheses, which were based on more data, mostly due to the placement of the root of the *Anelosimus* tree. Therefore, preliminary alignments were followed by manual and automated alignments taking into consideration the implied secondary structure of ITS2 (Fig. 1). I used the ITS2 database (online at http://its2.bioapps.biozentrum.uni-wuerzburg.de/cgi-bin/index.pl?about) to annotate the sequences and find the 5.8 and 28S flanking regions. *Anelosimus* ITS2 sequences were generally short, ranging from 223–305 bp. Non-ITS2 sequences were then removed and ITS2 secondary structure implied using the 4Sale software (Seibel et al. 2006, 2008). A standard model of ITS2 secondary structure was developed by Sehultz et al. (2005). This model has one long arm, or 'helix' (helix III), and three shorter helices (helices I, II, and IV), all four radiating from an area of a large loop. Secondary structure analyses reveal that the region that aligns poorly using Clustal corresponds to helix IV of the consensus ITS2 secondary structure model (Schultz et al. 2005) which is present in *Latrodectus*, *A. rupununi* Levi 1956, and a short version of it in *Enoplognatha*, *A. nigrescens* (Keyserling 1884) and *A. ethicus* (Keyserling 1884), but lost in the '*eximius* lineage' (Fig. 1).

Once this helix is identified, manual alignment of this region is facilitated, essentially aligning apparently homologous regions of helix IV in those taxa that have it, and inserting a gap for the entire arm region in those taxa that lack it. Automated alignment was also

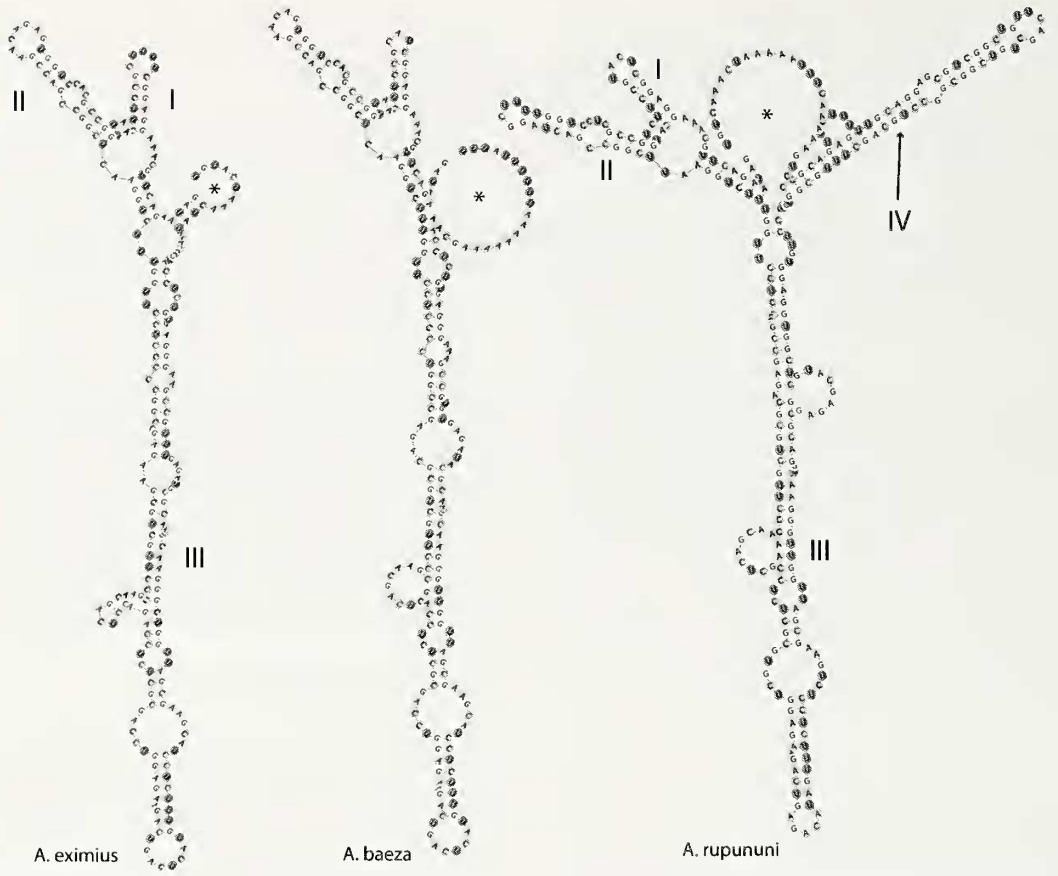

Figure 1.—Secondary structure (Bruccoleri layout) of ITS2 as implied by 4SALE for two species of the *eximius* lineage (*A. eximius* and *A. baeza*) and *A. rupununi*. The overall similar secondary structure reflects sequence similarity across most of the ITS2 sequence in these taxa that is readily alignable. However, *A. rupununi* has a helix (arrow), corresponding to helix IV of the ITS2 consensus structure of Schultz et al. (2005) that has been lost in the *eximius* lineage. A second region of difficult alignment is a loop region preceding this helix IV (stars).

conducted with the 4Sale software, using the remote 4Sale option. The automated alignment was not modified other than by fixing the first eight aligned characters, representing a five base pair sequence identical in all taxa, which had been rather randomly spread out. The remainder of the automated alignment did not contain conspicuous areas of misalignment. Aligned matrices and results are available from the author upon request.

The appropriate substitution model was selected with Modeltest (Posada and Crandall 1998), using the AIC criterion (Posada and Buckley 2004) with a parsimony tree chosen as the basis for Modeltest. The best model was GTR + Γ + I (Yang 1994). Bayesian analysis was performed using MrBayes V3.1.2 (Huelsenbeck and Ronquist 2001). The Markov chain Monte Carlo was run with four chains for 10,000,000 generations (repeated twice), sampling the Markov chain every 1000 generations, and the sample points of the first 5,000,000 generations were discarded as "burnin". Maximum likelihood analyses were conducted in the program Garli (Zwickl 2006), using the GTR + Γ + I model and 200 search replications. Parsimony analysis was done using TNT default settings under traditional search, with 1000 search replications. To calculate divergences among and within species in previous studies (Table 1), I downloaded sequences from GenBank via Mesquite (Maddison and Maddison 2009) and calculated uncorrected genetic distances in Mesquite.

The phylogenies are largely congruent using the Bayesian, likelihood, or parsimony criteria, and whether based on the manual or automated alignment (Fig. 2); hence, only the Bayesian results are discussed. To the extent that the current results are comparable to prior studies that included more taxa, they recapitulate the deeper-level phylogeny of Agnarsson et al. (2007, 2010) based on six molecular loci combined with morphology (Fig. 2). However, the analysis does not resolve relationships among closely related species of the *studiosus/jucuudus* groups and does not reflect strong mitochondrial population structure within *A. eximius*.

Within *Anelosimus*, therefore, the utility of ITS2 seems very limited at the lowest taxonomic level (within species, between closely related species), but higher at intermediate taxonomic levels. Other studies of closely related theridiid species have also found little to no informative variation among closely related species (Zhang et al 2004, Vink et al. 2008). However, in cases where population structuring is particularly strong, such as in trapdoor spiders (Bond & Stockman 2008) and cave-dwelling nesticids (Hedin 1997), ITS2 was found to be useful at the interspecific, and even intraspecific level. Based on this and prior studies, ITS2 is a useful and readily obtainable marker for phylogenetic studies that look at relationships within genera and families of spiders. In general, intraspecific variation is low in spiders (about 1% on average, Table 1), as is the variation between sister species, but the variation differs across groups and is notably high in some trapdoor and cave dwelling spiders (Table 1). Closely related *Anelosimus* and *Latrodectus* species have very low interspecific variation (typically < 1%, about 0.7% in *A. eximius*, which shows high mitochondrial variation), insufficient to resolve relationships among closely related species, or to diagnose species. The utility of ITS2 at lower taxonomic levels thus will vary depending on the group. At higher taxonomic levels the main difficulty will be extreme sequence divergence (e.g., 27% between *A. rupununi* and *A. eximius*), thus complicating alignment. However, ITS2 secondary structure can facilitate alignment of divergent

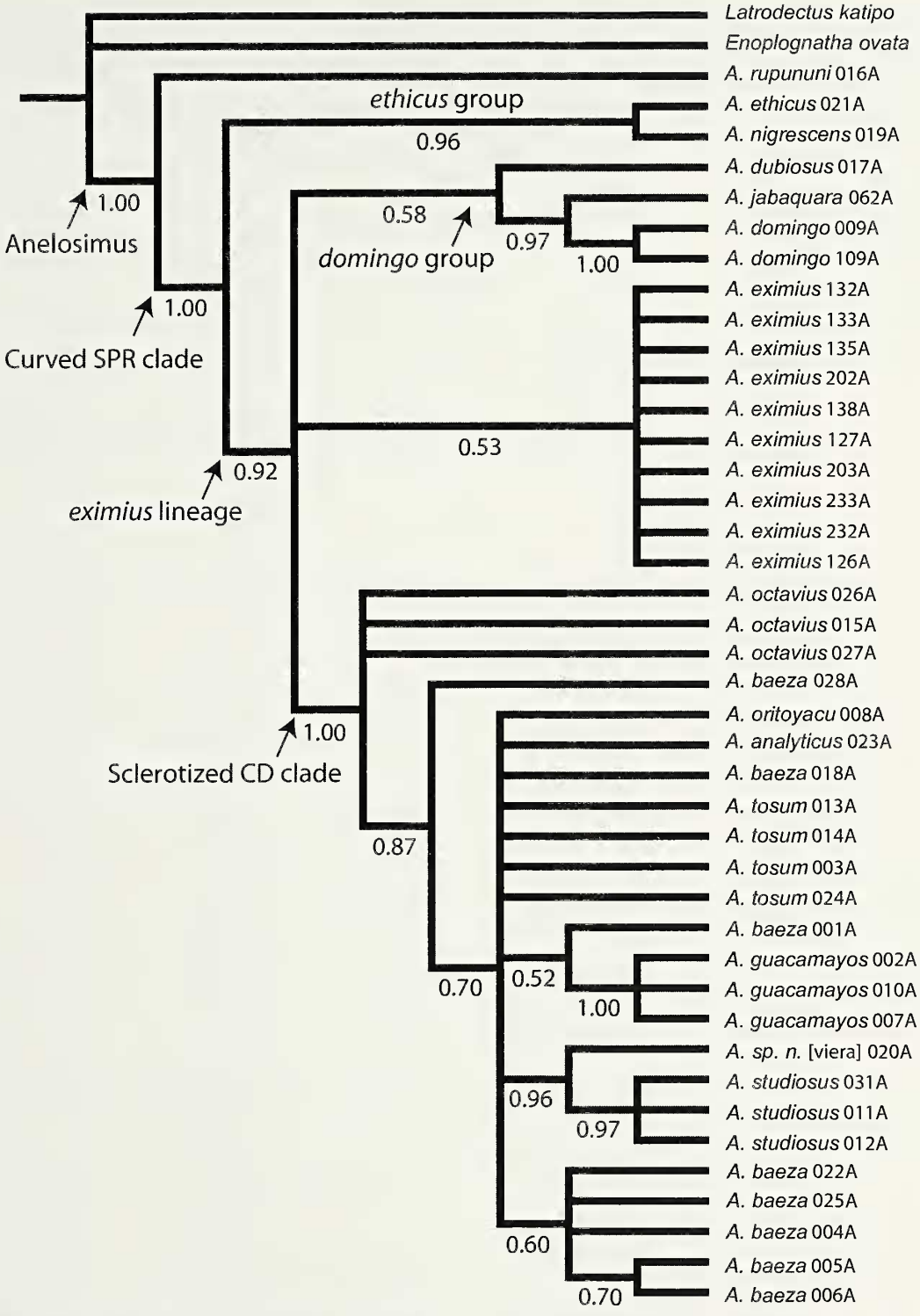

Figure 2.—50% Majority consensus from the Bayesian analysis of secondary-structure-informed manual alignment with numbers showing posterior probability support values. Major deeper level clades and species groups well supported by prior work are recovered: see clade labels. All labeled clades were recovered in all analyses, except the '*domingo* group'. However, at lower taxonomic levels very little variation was observed, resulting in low resolution. Within the 'sclerotized CD clade' (*A. analyticus* plus the *jucundus/studiosus* complex) only *A. guacamayos* Agnarsson 2006 and *A. studiosus* (Hentz 1850) were recovered as monophyletic, the relationships among species were largely unresolved and inconsistent with prior work (Agnarsson 2006, 2010; Agnarsson et al. 2007). Within *A. eximius*, a species showing population division and strong mitochondrial structuring, no phylogenetic structure was recovered.

Table 1.—ITS2 maximum intraspecific sequence divergences, and estimation of maximium and minimum divergences between sister species, in previously published studies of spiders. Estimated intraspecific sequence divergence is likely conservative overall, as some species were sampled only by 2–3 individuals. However, even for species sampled by 10 or more individuals and from geographically distant localities (e.g., *Latrodectus katipo*, *Anelosimus eximius*) the divergences were low.

Family	Genus	Species or putative species	Maximum intraspecific sequence divergence	Reference
Araneidae	*Poltys*	*illepidus*	0	Smith 2006
Araneidae	*Poltys*	*stygius*	0	Smith 2006
Araneidae	*Poltys*	*laciniosus*	0	Smith 2006
Cyrtaucheniidae	*Aptostichus*	clade 5	0.039	Bond and Stockman 2008[4]
Cyrtaucheniidae	*Aptostichus*	Clade 2	0	Bond and Stockman 2008
Cyrtaucheniidae	*Aptostichus*	Clade 3	0.025	Bond and Stockman 2008
Cyrtaucheniidae	*Aptostichus*	Clade 1	0.004	Bond and Stockman 2008
Linyphiidae	*Orsonwelles*	*graphica*	0.005	Hormiga et al. 2003
Linyphiidae	*Orsonwelles*	*macbeth*	0.01	Hormiga et al. 2003
Linyphiidae	*Orsonwelles*	*falstaffius*	0	Hormiga et al. 2003
Linyphiidae	*Orsonwelles*	*polites*	0	Hormiga et al. 2003
Linyphiidae	*Orsonwelles*	*otliello*	0.005	Hormiga et al. 2003
Linyphiidae	*Orsonwelles*	*anbersonorum*	0	Hormiga et al. 2003
Linyphiidae	*Orsonwelles*	*malus*	0.002	Hormiga et al. 2003
Linyphiidae	*Orsonwelles*	*arcanus*	0.02	Hormiga et al. 2003
Linyphiidae	*Orsonwelles*	*calx*	0.002	Hormiga et al. 2003
Linyphiidae	*Orsonwelles*	*ventus*	0.02	Hormiga et al. 2003
Lycosidae	*Pardosa*	*astigera*	0.03	Chang et al. 2007
Lycosidae	*Pardosa*	*astigera* (phenotype A)	0.003	Chang et al. 2007
Lycosidae	*Pardosa*	*astigera* (phenotype B)	0.005	Chang et al. 2007
Nesticidae	*Nesticus*	*barri*	0.0025	Hedin 1997[1]
Nesticidae	*Nesticus*	*barrowsi*	0.0102	Hedin 1997
Nesticidae	*Nesticus*	*bishopi*	0.0051	Hedin 1997
Nesticidae	*Nesticus*	*cooperi*	0.0059	Hedin 1997
Nesticidae	*Nesticus*	"*dellingeri*"	0.0076	Hedin 1997
Nesticidae	*Nesticus*	*gertschi*	0.0152	Hedin 1997
Nesticidae	*Nesticus*	*minus*	0.0119	Hedin 1997
Nesticidae	*Nesticus*	*nasicus*	0.0077	Hedin 1997
Nesticidae	*Nesticus*	*silvanus*	0.0034	Hedin 1997
Nesticidae	*Nesticus*	*stupkai*	0.0102	Hedin 1997
Nesticidae	*Nesticus*	*carteri*[2]	0.0321	Hedin 1997
Nesticidae	*Nesticus*	nov. sp	0.0051	Hedin 1997
Nesticidae	*Nesticus*	*paynei*	0.0076	Hedin 1997
Nesticidae	*Nesticus*	*tennesseensis*	0.0085	Hedin 1997
Salticidae	*Havaika*	OK9, OK24, OW28, OW29	0.0176	Arnedo and Gillespie 2006[3]
Salticidae	*Havaika*	OK8, OK23, OW111, OW158	0.03	Arnedo and Gillespie 2006[3]
Salticidae	*Havaika*	H83, H109, H137	0.005	Arnedo and Gillespie 2006[3]
Salticidae	*Havaika*	EM128, WM88, WM159	0.018	Arnedo and Gillespie 2006[3]
Salticidae	*Havaika*	K85, K86, K87	0.03	Arnedo and Gillespie 2006[3]
Salticidae	*Havaika*	EM81, MK82, WM89	0.0025	Arnedo and Gillespie 2006[3]
Salticidae	*Havaika*	H10, H110, EM90	0.015	Arnedo and Gillespie 2006[3]
Salticidae	*Havaika*	WM88, WM159, EM128	0.018	Arnedo and Gillespie 2006[3]
Theridiidae	*Latrodectus*	*katipo*[5]	0.002	Vink et al. 2008
Theridiidae	*Latrodectus*	*hasselti*	0	Vink et al. 2008
Theridiidae	*Latrodectus*	*hasselti*	0.0027	Zhang et al. 2004
Theridiidae	*Latrodectus*	*mactans*[6]	0.014	Zhang et al. 2004
Theridiidae	*Anelosimus*	*eximius*	0.007	This study
Theridiidae	*Anelosimus*	*domingo*	0	This study
Theridiidae	*Anelosimus*	*tosum*	0.008	This study
Theridiidae	*Anelosimus*	*studiosus*	0.01	This study
Theridiidae	*Anelosimus*	*guacamayos*	0.002	This study
Theridiidae	*Anelosimus*	*octavius*	0.007	This study
Theridiidae	*Anelosimus*	*baeza*	0.02	This study
Average			**0.01**	

Table 1.—Continued.

	Interspecific sequence divergence (sister species)		
	Min	Max	Reference
Nesticus	0.40%	~9%	Hedin 1997
Latrodectus	0%	0.50%	Vink et al. 2008
Latrodectus	0%	0.83%	Zhang et al. 2004
Havaika	2%	4%	Arnedo and Gillespie 2006[3]
Pardosa	2.50%	6.70%	Chang et al. 2007
Orsonwelles	0.70%	5.90%	Hormiga et al. 2007
Anelosimus	0.60%	2.80%	This study
Poltys	0.70%	~10%	Smith 2006
Aptostichus	2.20%	5.30%	Bond and Stockman 2008

[1] Note that multiple individuals within populations always had zero sequence divergence, interspecific sequence divergences reflect those among isolated populations
[2] Represented two species, each with intraspecific divergence less than 1.5%
[3] Informal species, reflecting putative species from Fig. 5 in Arnedo and Gillespie (2006), codes in 'species' column refer to specimens
[4] Sequences from 'clade 4' were not found on Genbank
[5] One variable site
[6] More divergence found within than between individuals

sequences (Young & Coleman 2004) (Fig. 1). Based on my findings and those of Vink et al. (2008) and Zhang et al. (2004), ITS2 does not emerge as a suitable choice of universal DNA barcode.

LITERATURE CITED

Agnarsson, I. 2006. A revision of the New World *eximius* lineage of *Anelosimus* (Araneae, Theridiidae) and a phylogenetic analysis using worldwide exemplars. Zoological Journal of the Linnean Society 146:453–593.

Agnarsson, I., L. Avilés, J.A. Coddington & W.P. Maddison. 2007. The phylogeny of the social *Anelosimus* spiders (Araneae: Theridiidae) inferred from six molecular loci and morphology. Molecular Phylogenetics and Evolution 43:833–851.

Agnarsson, I., M. Kuntner, J. Coddington & T.A. Blackledge. 2010. Shifting continents, not behaviours: independent colonization of solitary and subsocial *Anelosimus* spider lineages on Madagascar (Araneae, Theridiidae). Zoologica Scripta 39:75–87.

Alverez, I. & J.R. Wendel. 2003. Ribosomal ITS sequenees and plant phylogenetic inference. Molecular Phylogenetics and Evolution 29:417–434.

Arnedo, M.A. & R.G. Gillespie. 2006. Species diversification patterns in the Polynesian jumping spider genus *Havaika* Prószyński, 2001 (Araneae, Salticidae). Molecular Phylogenetics and Evolution 41:472–495.

Avilés, L. 1997. Causes and consequences of cooperation and permanent-sociality in spiders. Pp. 476–498. *In* The Evolution of Social Insects and Arachnids. (J.C. Choe & B.J. Crespi, eds.). Cambridge University Press, Cambridge, UK.

Bailey, C.D., T.G. Carr, S.A. Harris & C.E. Hughes. 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Molecular Phylogenetics and Evolution 29:435–455.

Bond, J.E. & A. Stockman. 2008. An integrative method for delimiting cohesion species: finding the population-species interface in a group of Californian trapdoor spiders with extreme genetic divergence and geographic structuring. Systematic Biology 57:628–646.

Ben-David, T., S. Melamed, U. Gerson & S. Morin. 2007. ITS2 sequences as barcodes for identifying and analyzing spider mites (Acari: Tetranychidae). Experimental and Applied Acarology 41:169–181.

Chang, J., D. Song & K. Zhou. 2007. Incongruous nuclear and mitochondrial phylogeographie patterns in two sympatric lineages of the wolf spider *Pardosa astrigera* (Araneae: Lycosidae) from China. Molecular Phylogenetics and Evolution 42:104–121.

Coleman, A.W. 2009. Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. Molecular Phylogenetics and Evolution 50:197–203.

Hedin, M.C. 1997. Speciation history in a diverse clade of habitat specialized spiders (Araneae: Nesticidae: *Nesticus*): inferences from geographic-based sampling. Evolution 51:1929–1945.

Hormiga, G., M. Arnedo & R.G. Gillespie. 2003. Speciation on a conveyor belt: sequential colonization of the Hawaiian Islands by *Orsonwelles* spiders (Araneae, Linyphiidae). Systematic Biology 52:70–88.

Huelsenbeck, J.P. & F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Maddison, W.P. & D.R. Maddison. 2009. Mesquite: a modular system for evolutionary analysis. Version 2.72. Online at http://mesquiteproject.org.

Park, M.H., C.J. Sim, J. Baek & G.S. Min. 2007. Identification of genes suitable for DNA barcoding of morphologically indistinguishable Korean Halichondriidae sponges. Molecules and Cells 23:220–227.

Posada, D. & K.A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818.

Posada, D. & T.R. Buckley. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology 53:793–808.

Seibel, P.N., T. Müller, T. Dandekar & M. Wolf. 2008. Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. BMC Research Notes 1:91.

Schultz, J., S. Maisel, D. Gerlach, T. Müller & M. Wolf. 2005. A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. RNA 11:361–364.

Schultz, J. & M. Wolf. 2009. ITS2 sequence–structure analysis in phylogenetics: A how-to manual for molecular systematics. Molecular Phylogenetics and Evolution 52:520–523.

Seibel, P.N., T. Müller, T. Dandekar, J. Schultz & M. Wolf. 2006. 4SALE - A tool for synchronous RNA sequence and secondary structure alignment and editing. BMC Bioinformatics 7:498.

Smith, H.M. 2006. A Revision of the Genus *Poltys* in Australasia (Araneae: Araneidae). Records of the Australian Museum 58:43–96.

Thompson, J.D., D.G. Higgins & T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence align-

ment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22:4673–4680.

Vink, C.J., P.J. Sirvid, J. Malumbres-Olarte, J.W. Griffiths, P. Paquin & A.M. Paterson. 2008. Species status and conservation issues of New Zealand's endemic *Latrodectus* spider species (Araneae: Theridiidae). Invertebrate Systematics 22:589–604.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. Journal of Molecular Evolution 39:306–314.

Young, I. & A.W. Coleman. 2004. The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a *Drosophila* example. Molecular Phylogenetics and Evolution 30:236–242.

Zhang, D., W.B. Cook & N.V. Horner. 2004. ITS2 rDNA variation of two black widow species, *Latrodectus mactans* and *Latrodectus hesperus* (Araneae, Theridiidae). Journal of Arachnology 32:349–352.

Zwickl, D.J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. Dissertation, The University of Texas at Austin, Texas.

Manuscript received 2 January 2010, revised 28 April 2010.